Sound Classification and Localization Based on Biology Hearing Models and Multiscale Vector Quantization

John S. Baras

Center for Auditory and Acoustic Research

Electrical and Computer Engineering Department

and the Institute for Systems Research

University of Maryland College Park

Center for Auditory and Acoustic Research

Motivation

- **Algorithms motivated by similar processing in animals and humans:**
  - **Hearing and sound classification**
  - **Vision and identification of objects**
- **Text-independent robust speaker identification**
  - **Identifying the speaker from the "music" of his voice**
- **Speaker-independent speech recognition**
  - Identifying phonemes, vowels, words from their inherent sounds
- **Identification of musical instruments ("timbre")**

Applications to acoustic signal recognition

- **Fault identification in tools and wear prediction**
- **Ground vehicle identification from array microphones**

NEXT CHALLENGE: Biology Inspired Sensor Network processing

Center for Auditory and Acoustic Research

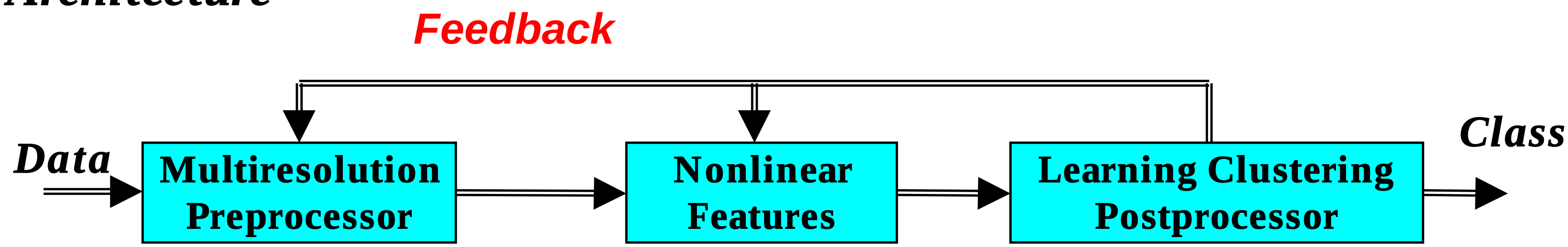
Acoustic Vehicle Classification Objectives and Challenges

- Develop systematic methodologies and algorithms; not *ad hoc*
- Robust Target ID (wrt environment, terrain, speed)
- Algorithms for combined DOA (localization) and target ID
  - Localization assisted ID
  - ID assisted localization
- Multi-target detection, ID and DOA; separation of closely spaced targets
- Robust feature extraction from auditory models; dynamic DOA and ID
- Algorithm evaluation in the field and comparison against conventional algorithms for detection, DOA and ID

Center for Auditory and Acoustic Research

Multiresolution Adaptive Acoustic Classification

- ***Architecture***

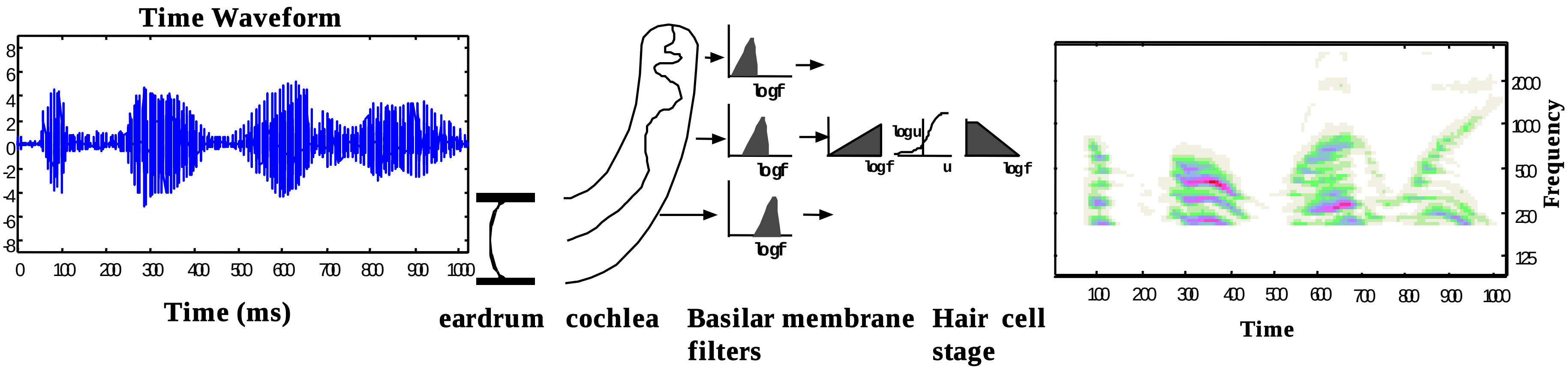


- **Architecture and formulation address two most important issues:**
  - **Progressive classification; Which features to use and when**
  - **Efficient design of databases for reference signals and fast search**
- **Trade-off between efficiency in features (compression) and accuracy in classification leads to**
- **Mathematical formulation of the problem:**
  - **Combined compression and classification for general signals**
  - **Content-based feature extraction and use for classification**

Center for Auditory and Acoustic Research

Multiresolution Preprocessor: Auditory Filtering

Two auditory filters, motivated and designed according to acoustic physiology and acoustic cortex models, were used to compute the timbre spectrogram of one particular subframe in each frame



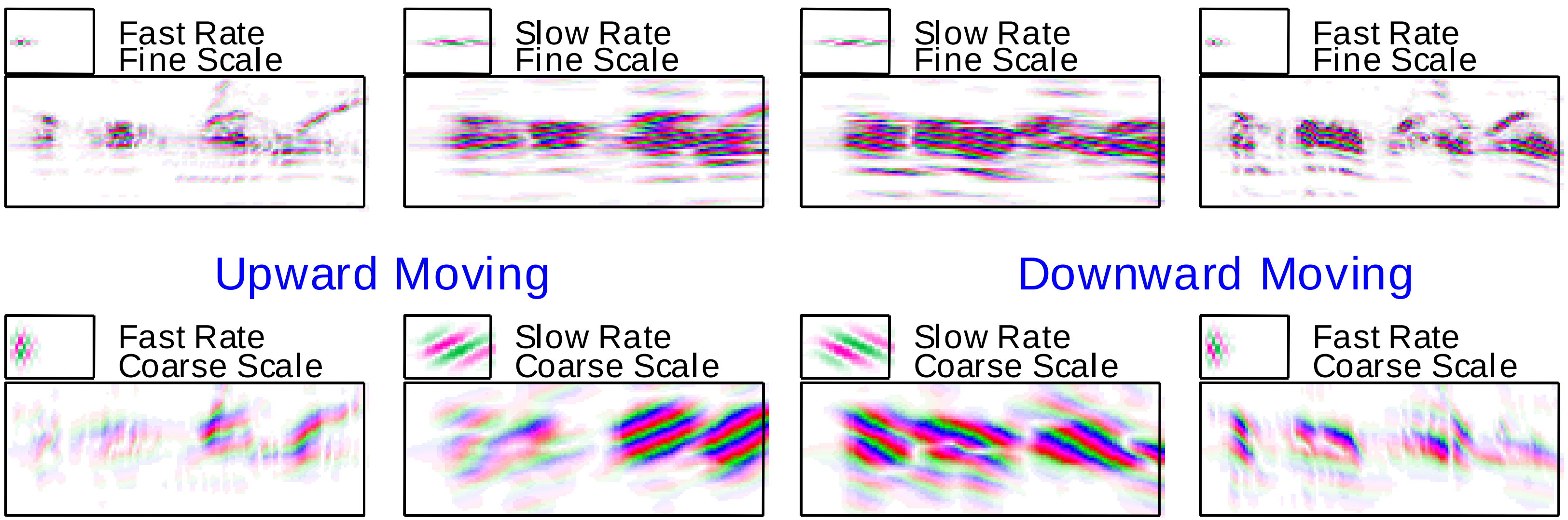
- The first filter mimics the action of the inner ear
- Computes the spectrogram of the sound sample, and performs various nonlinear operations, which models the nonlinear fluid-cilia couplings and ionic channels of conduction

(Wavelet Transform)

Center for Auditory and Acoustic Research

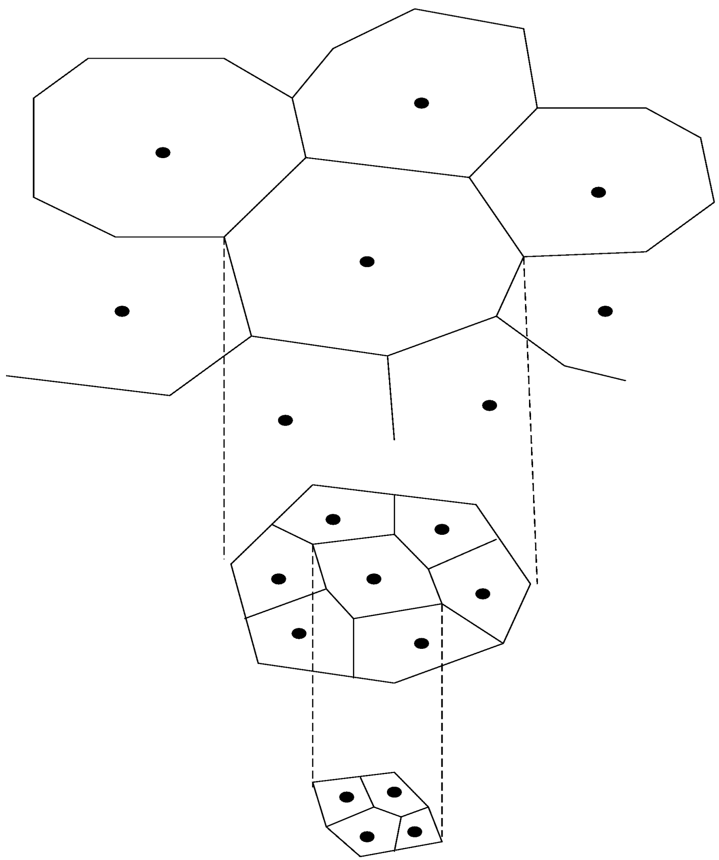
Multiresolution Preprocessor: Auditory Filtering

Multiresolution cortical filter outputs



- **The second filter models the multiscale processing of the signal that happens in the auditory cortex**
- **A Ripple Analysis Model, using a ripple filter bank, acts on the output of the inner ear to give multiscale spectra of the sound timbre (Wavelet Transform)**

Postprocessor: Multi Resolution (Wavelet) Tree Structured Vector Quantization (WTSVQ)



- First perform a multiresolution wavelet representation of the signals
- Consider each signal f at different resolutions

  $S^0 f, \; S^1 f, \ldots, \; S^{J^*} f$
- Proceed by partitioning the signal space at various resolutions in progressively finer cells
- *Greedy algorithm* works by splitting the cell with maximum distortion using finer resolution data

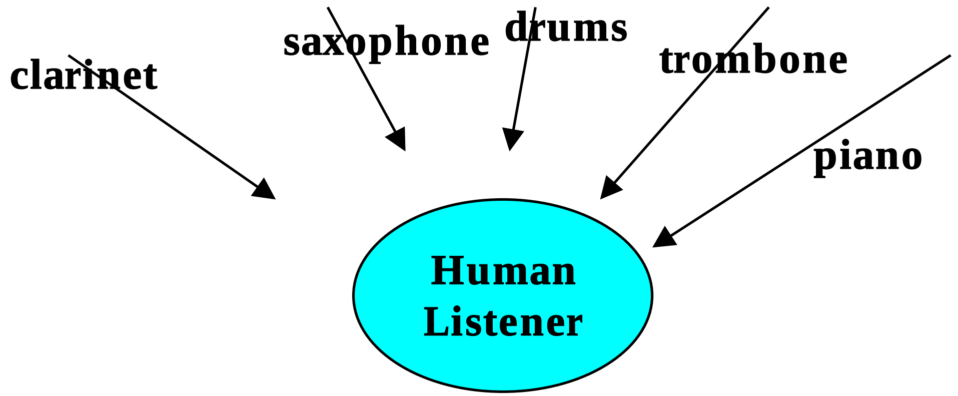
Layer in tree $l = J^* - m$, m the scale (top layer 0: coarsest)

Cell labels: (layer, index) or (scale, index)

Center for Auditory and Acoustic Research

Approach

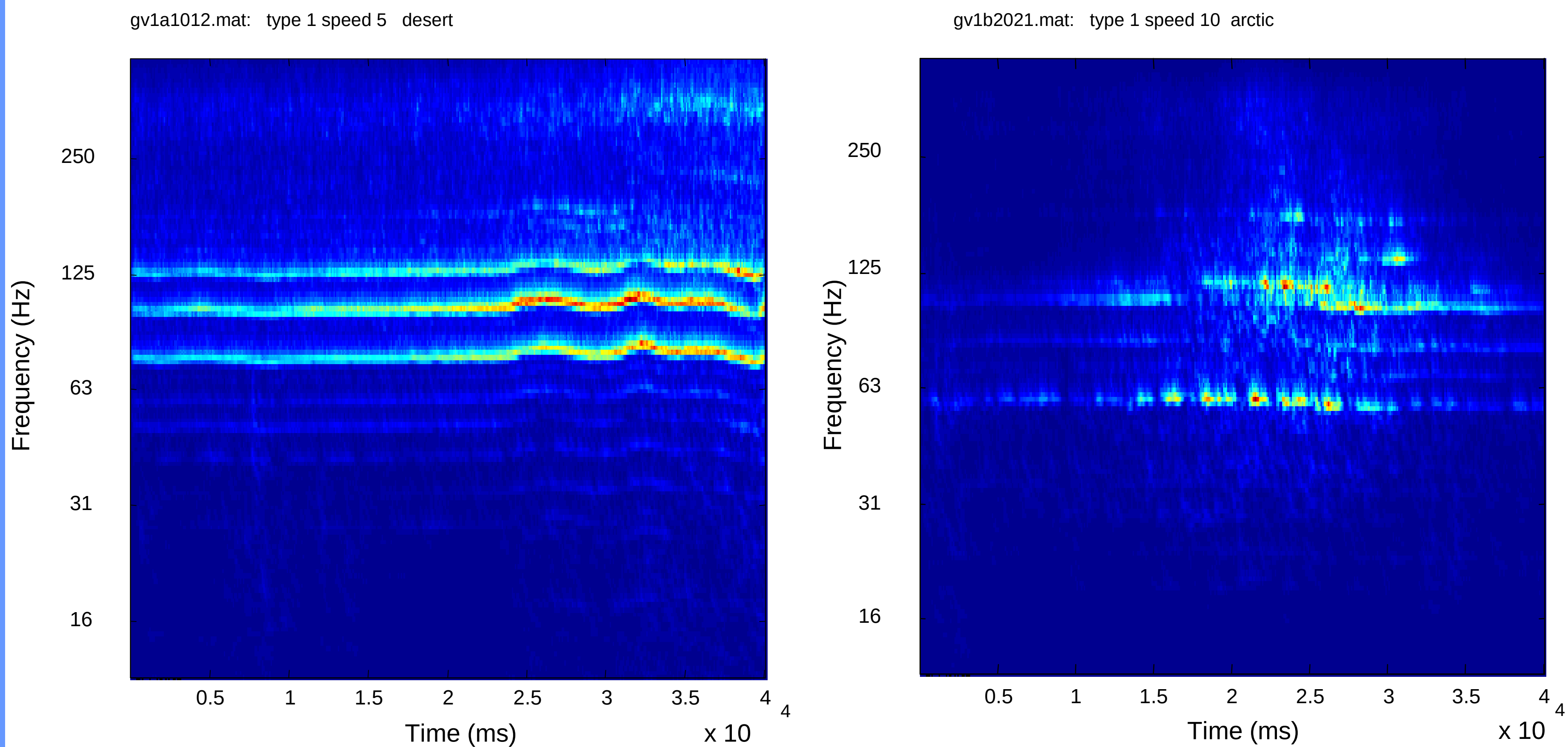
Can we mimic and understand the ability of humans to do partial recognition of musical instruments and DOA in a combined and mutually enhancing fashion?



- Combine the Stereausis model and its derivatives , with the Auditory filtering multiscale VQ algorithms
- Using the cochlea, cortical, or combined spectra, perform DOA on a “per frequency band basis”
- Combine portions of spectra according to DOA
- Use the multiscale classifier to ID portions of spectra tagged by angle, as compared to stored vehicle spectra
- Repeat the cycle as the scenario evolves

Center for Auditory and Acoustic Research

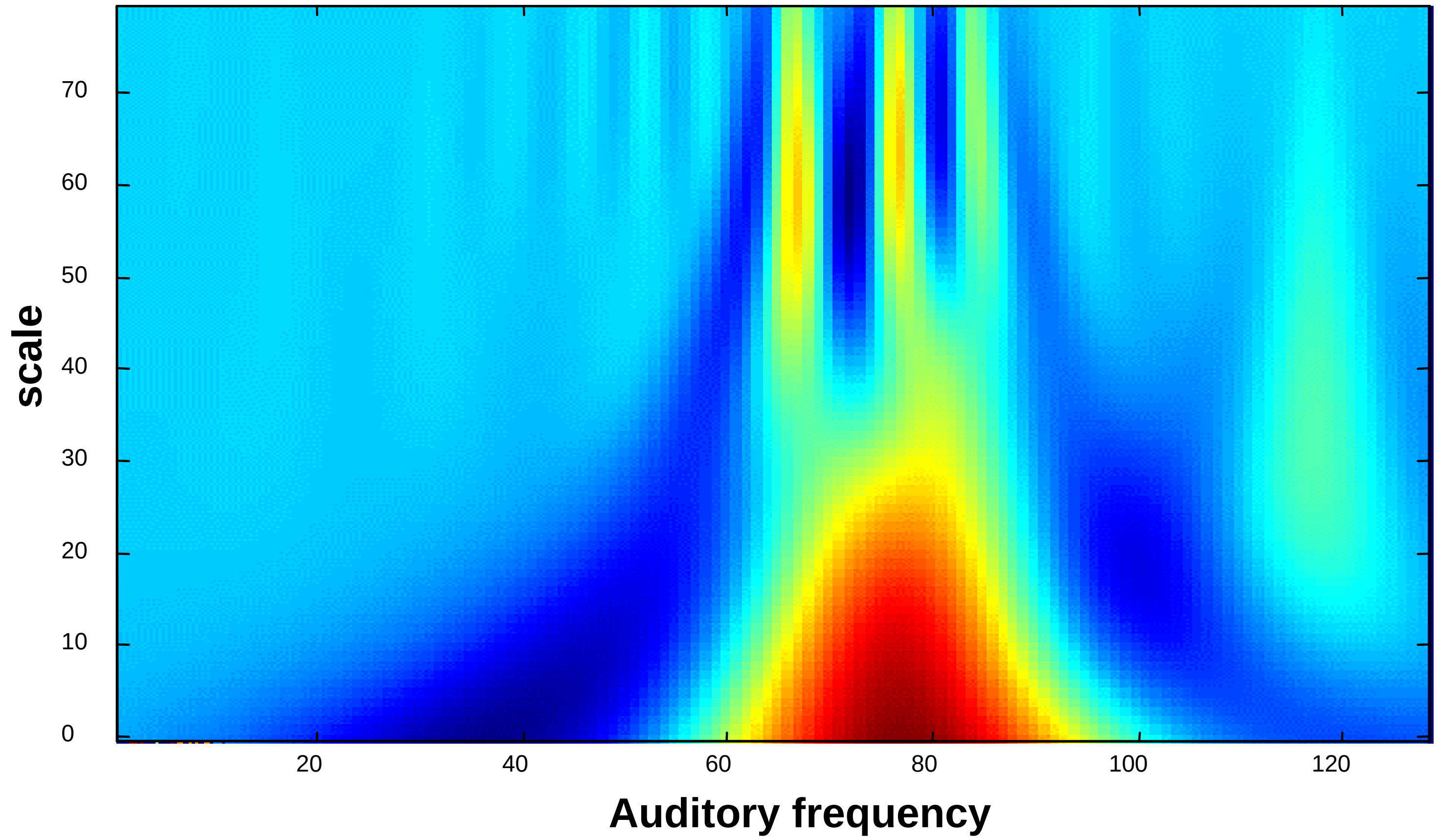
Auditory Processing of Vehicle Acoustic Signals: Cochlea



Auditory processing for vehicle signals (cochlear filter banks)
Left: vehicle type 1, speed 5km/hr. Right: vehicle type 1, speed 10km/hr

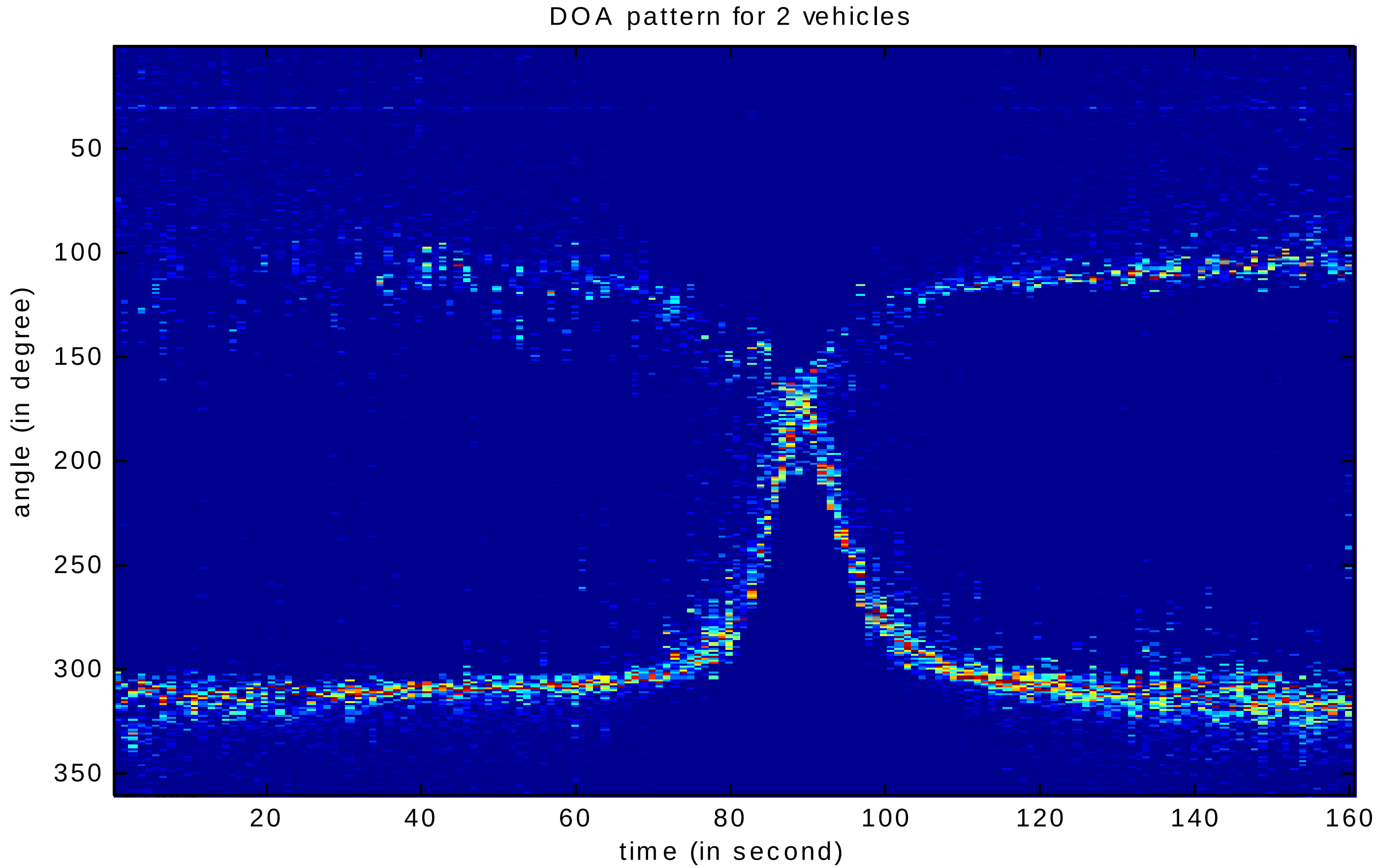
Center for Auditory and Acoustic Research

Auditory Processing of Vehicle Acoustic Signals: Cortex



Example of multi-resolution representation from cortical module

Stereausis Output for Two Vehicles



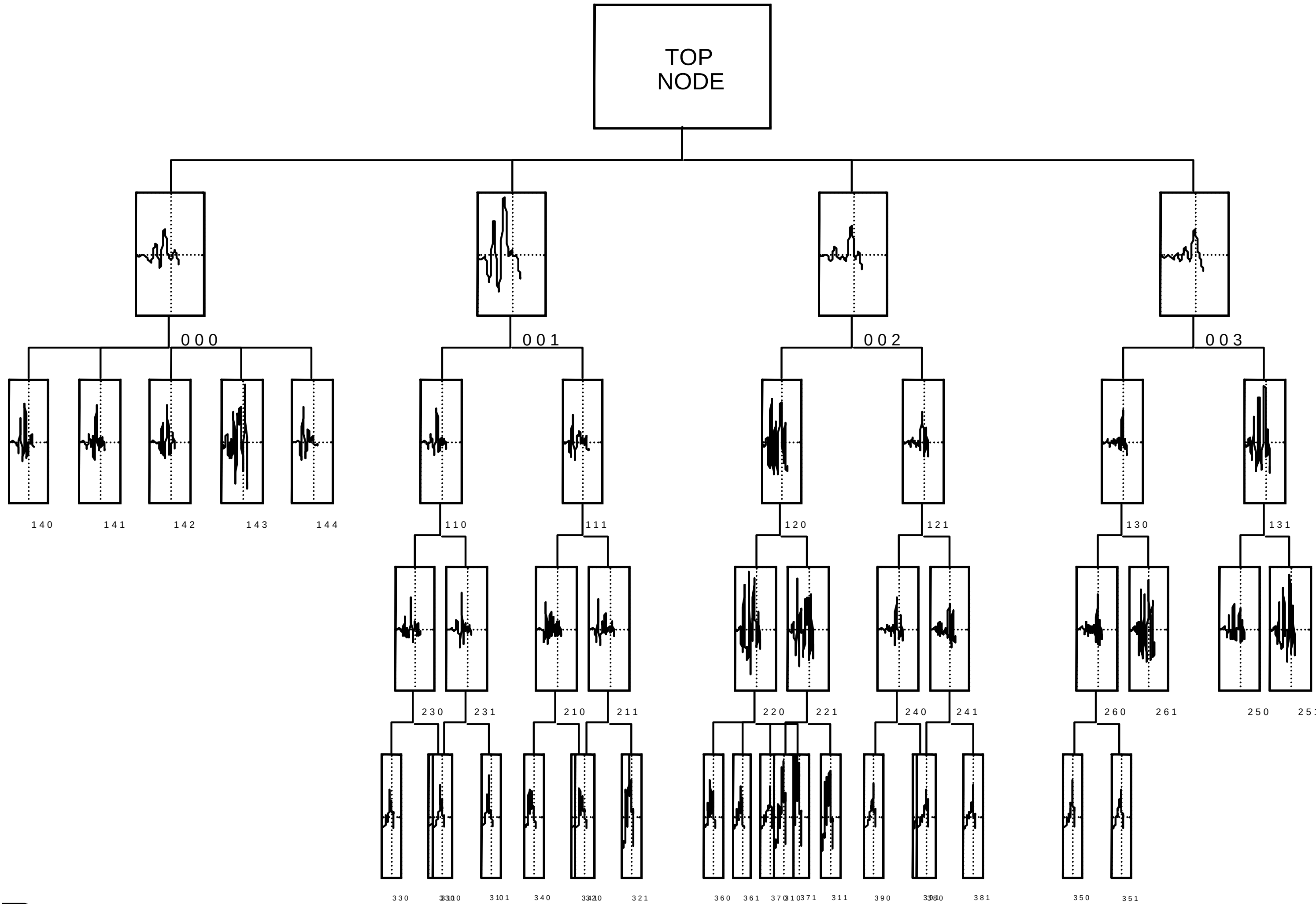
Relatively easy case: Large angular separation between two vehicles

Center for Auditory and Acoustic Research

Leaf Node Entropies for PTSVQ Tree of Vehicle Type 8

cell entropy

1 4 0	1.3570
1 4 1	0.9503
1 4 2	1.1779
1 4 3	1.0735
1 4 4	1.3022
2 5 0	0.6365
2 6 1	0
3 1 0	0.5765
3 1 1	0.2993
3 2 0	0.7516
3 2 1	0.4765
3 3 0	0.7633
3 3 1	0.5670
3 4 0	0.4540
3 4 1	0.4384
3 5 0	0.2728
3 5 1	0.4975
3 6 0	0.5313
3 6 1	0.3061
3 7 0	0.6054
3 7 1	0.6383
3 8 0	0.4824
3 8 1	0.5377
3 9 0	0.5044
3 9 1	1.2556
3 10 0	1.0144
3 10 1	1.1967

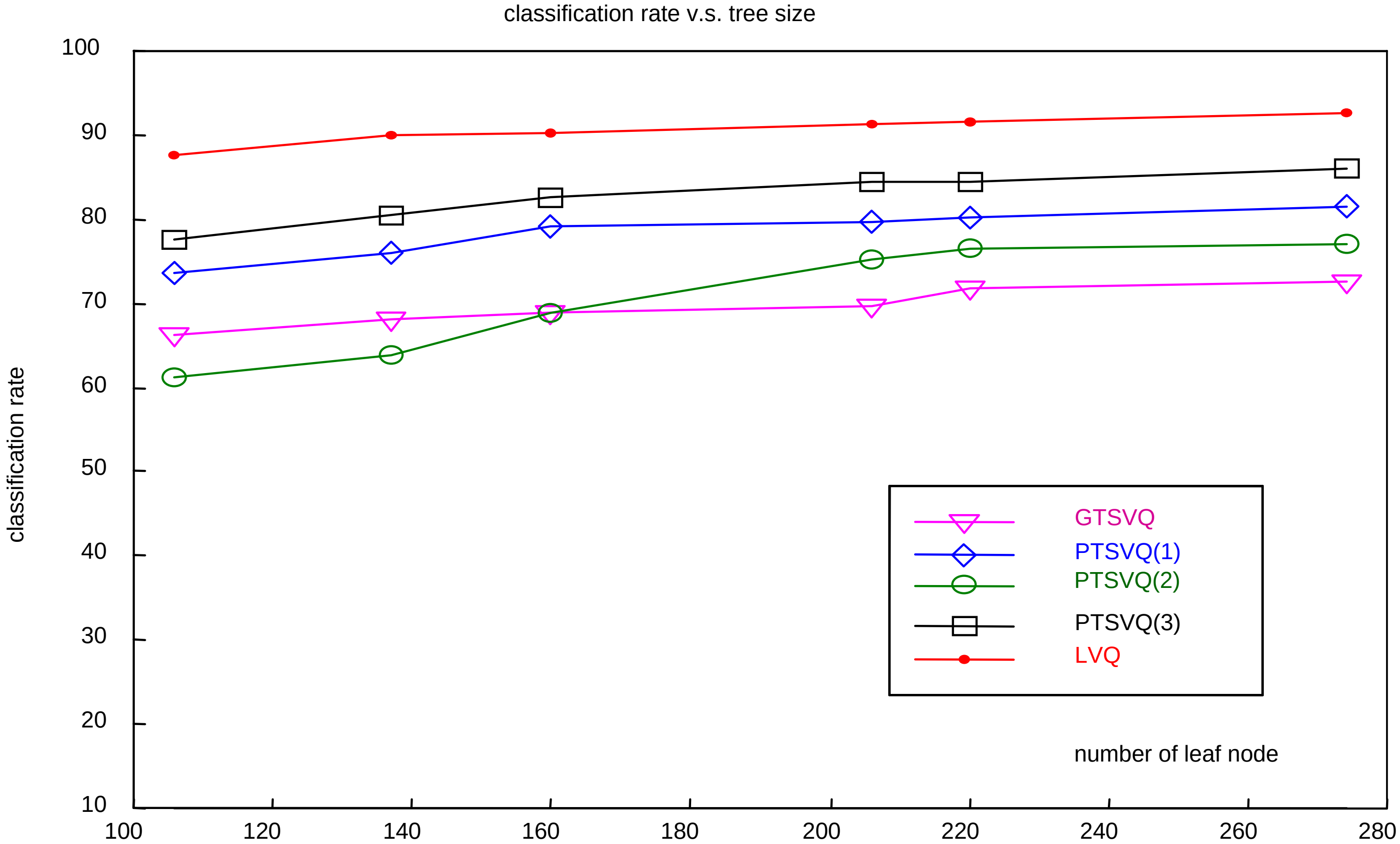


Center for Auditory and Acoustic Research

Options in Applying WTSVQ to Acoustic Vehicle Classification

- **GTSVQ: A global tree-structured multi-resolution clustering mechanism that mimics the aggressive and topological hearing capabilities of biological systems. Here a global tree is built on training data from all vehicles. New vehicle insertion problem.**
- **LVQ: A supervised learning neural network, LVQ achieves optimal classification in the Bayes sense. It has the disadvantages of a long search time and sensitivity to initial conditions.**
- **Parallel TSVQ (PTSVQ): build one (or more) trees for each vehicle. It achieves a trade-off between GTSVQ and LVQ on classification performance and search time. Easy new vehicle insertion.**
- **The following node allocation schemes are examined for PTSVQ:**
  - **PTSVQ(1): Allocation based on sample a priori probability**
  - **PTSVQ(2): Allocation based on equal distortion**
  - **PTSVQ(3): Allocation according to vehicle speed**

Performance Comparisons among Options

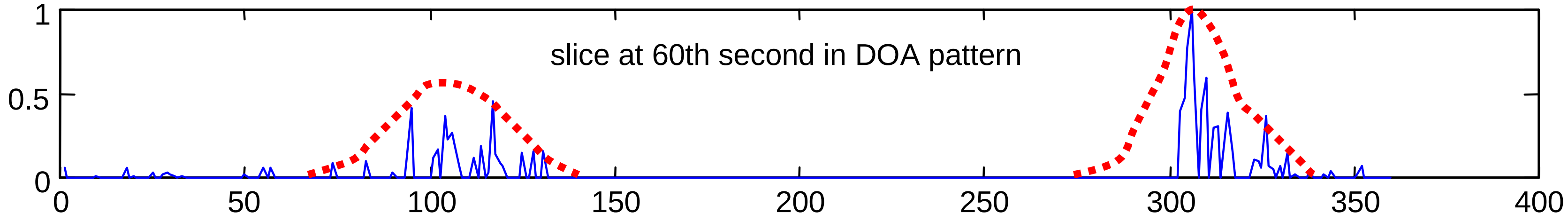


Classification Performance: 70% samples for training, 30% for testing (same microphone)

Center for Auditory and Acoustic Research

Tagging Portions of Spectra Based on "per Band" DOA Estimates

- Angular position of ach peak corresponds to DOA estimate from each cochlea band
- Can use up to 128 bands
- Amplitude indicates signal energy in the band



- Low pass filtering is performed on groups of band amplitudes and the resulting peak is used as the DOA estimate for the vehicle
- Cluster according to angular position of peaks: spectral portions tagged by angle